PACIFIC SOUTHWEST
Forest and Range
Experiment Station

FOREST SERVICE
U.S.DEPARTMENT OF AGRICULTURE
P.O. BOX 245, BERKELEY, CALIFORNIA 94701

FAMULUS: a computer-based system for augmenting personal documentation efforts

Hilary D. Burton Robert M. Russell Theodor B. Yerke

U.S.D.A. Forest Service
Research Note PSW-193
1969

ABSTRACT: FAMULUS is a computer-based system designed to support the documentation activities of the individual scientist with minimum interference in his information-organizing habits and instincts. It operates economically on large, third-generation computers. For the first time the user is offered editing, sorting, indexing, vocabulary-building, searching, and file-revision features in a package which leaves him free to structure his input according to his idiosyncratic needs. Existing literature collections on FAMULUS are discussed to show the system's flexibility.

RETRIEVAL TERMS: FAMULUS; information system design; personal documentation; information retrieval; computer programs–FAMULUS.
OXFORD: U002–U681.3:U634.0.

Many systems analysts and research administrators regard with great skepticism the personal documentation systems which most scientists maintain. These systems are viewed as evidence that the formal bibliographic services provided by the researcher's supporting institution are poorly designed or functioning improperly. Accordingly, attention has been focused chiefly on improvement of library services and related bibliographic systems.

It is not yet widely understood that personal information systems are necessary to researchers because their orientation is basically antithetical to that of the formal system. The formal system is designed to be objective, inclusive, and normative. It may also tend to become static. The personal system is by definition subjective, heuristic, and dynamic. It provides its user with an information transfer environment dominated by his own viewpoint. He is the favored observer of his information universe. He determines what goes into the system and the relevance of retrieval results. And in this kind of documentation "relevance" is based on ever-varying subjective criteria.

Properly understood, both the formal and the personal systems contribute to the satisfaction of information needs by providing complementary services. To ignore either of the two systems, or to force one to substitute for the other, results in higher cost to the user and a lowering of the quality of over-all service.

Forest Service exploration of computer-based personal documentation began in 1964 at the Pacific Southwest Forest and Range Experiment Station in Berkeley, California. A research scientist started

---

Prepared for presentation at the 32nd annual meeting, American Society for Information Science, San Francisco, Calif., October 1-4, 1969.

LIBRARY FILE COPY

consultations with the Station librarian on the possibility of using the library's author-title catalog program for the processing of his own citation collection. Like most such collections maintained by scientists, his had become unmanageable. He wanted a way to maintain an up-to-date inventory of his holdings and a means for retrieving sets of items which he would periodically specify. Initial analysis of his collection structure and of the parameters built into the library's computer program was favorable. Within 6 months several other scientists with similar problems arranged with the librarian to have their references keypunched for processing by CATALOG, the name of the library program at that time.

We worked for 2 years with our pioneer users and their personal collections, during which time the concept of personal documentation and its support requirements was greatly advanced. Actual experience with various users demonstrated their information-handling habits to be extremely idiosyncratic. These individual differences come from personal (psychological) characteristics and from inherent variations in research disciplines. The more we could free the user to be idiosyncratic, the more efficient he felt the system to be.

In 1966, a computer services unit was established in the Station library. Its original function was to aid users of Station-produced and other programs and to maintain and operate the library's programs. The unit very quickly also became a support service for the growing number of personal collections. The increase in such collections began to reveal the limitations of CATALOG for this work, and the *desiderata* for a new system became quickly apparent. Its guiding principle had to be the greatest possible editing, sorting, deleting, revising, indexing, and searching ease for the user, within a framework of feasible input and item-size limits. Support for this work as potentially a Forest Service-wide application was received from the Washington office. The Station's senior programer, Robert M. Russell, was given the programing assignment. Because of the system's inclusive support features for personal documentation, it was named FAMULUS.[1]

TYPES OF COLLECTIONS

The general outline of FAMULUS was completed early in 1967, and programing of a Fortran IV version for the CDC 6400 computer began in June. It was completed in June 1968. All existing computer-based collections were shifted from CATALOG to FAMULUS, using an option built into the latter system. No reformatting or restructuring burdens were placed on the users. During the programing year, we encouraged new users to prepare their input material. More than a dozen scientists had their collections keypunched, and FAMULUS began operation with a very diverse clientele whose collections were structured in quite different ways.

Rigorously-structured collections. – Several users devised elaborately-structured formats for their input, similar in philosophy to a MARC approach. The Conrad collection (Appendix I, *7*)[2] in the field of range management exemplifies the type. His input breaks the ten main fields of FAMULUS into sub-fields, allowing identification within each citation of methodology, readability, type of data included, literature cited, etc.

In-depth collections. – The Anderson (Appendix I, *1*) collection comprises about 3,000 citations in watershed hydrology. It is a major bibliographic collection in this very specialized field. Each citation includes keywords assigned from the title and the body of the item as well as personally developed retrieval terms. Anderson also includes a 200- to 400-word abstract of his own, summarizing the item's contents from his point of view. Messenger's file (Appendix I, *2*) is on ecological entomology. Although abstracts are not included, he has done extensive indexing of each citation.

Broad-based collections. – The Remote Sensing collection (Appendix I, *3*) typifies this group. Information included ranges from technical specifications for photographic equipment to NASA reports on thermal sensors. Each citation is assigned to a general category and a sub-category. When FAMULUS sorts the tape record on category designations, the user obtains a classified catalog to his collection. Other projects whose collections consist of information synthesized from a variety of ancillary disciplines belong in this group (Appendix I, *4,5,6*).

Multi-user collections. – The Crisp collection (Appendix I, *8*) is one of several files which are conjoined to form the multi-user collection of the Experiment Station's Insecticide Evaluation Project. In this highly interdisciplinary project, each member works on a different aspect of the over-all problem. Periodically, staff members search each other's files. To do this they must first discuss their indexing approaches and clarify how A uses term J in his file as different from B's use of the same term in his own file. These mutual explorations and search operations, the participants report, lead to heightened awareness and understanding of their colleagues' work. Instead of forcing each participant to a standard terminology,

APPENDIX II

FAMULUS sub-systems. A thumbnail rundown of the functions performed by each program:

EDIT transfers card input to tape, checking the citations for gross errors in the process. Error-free citations are written on the tape. During subsequent jobs, EDIT can make corrections or changes, delete citations, or add new ones. Changes may be made to single citations, groups of citations, or to the collection as a whole.

SORT allows the user to rearrange the fields within a citation in any order he wants. For instance a standard citation might include author, title, publisher, date of publication, and accession number, in that order. SORT could rearrange the fields to be accession number, author, date of publication, title, publisher, or into any desired order. All 10 fields may be reordered to any depth.

MERGE provides the user with updating capacity as well as ability to work in a modular approach. For instance, with large collections it is economically feasible to process sections of 5,000 or fewer citations and then merge the section into a master file.

GALLEY is the printing unit of FAMULUS. The original EDIT format citations are formatted to standard bibliographic form. The user may specify the print width and so obtain notebook- or binder-compatible printout rather than standard printout size. Width can be varied from 20 columns to 128 columns.

OSSIFY punches card deck equivalents of tape files. Primarily, it has two uses: (1) to provide a safety deck in case the tape should be damaged; (2) to use where massive corrections must be made and it is preferable to make them on the card citations directly.

As the name implies, **INDEX** provides the user with a list of the citations indexed under keyword terms. The terms, which are user supplied, may be single words or word phrases. Up to 500 citations can be indexed in one run. For larger files, the partial indexes can be merged.

VOCAB builds vocabulary records for any user-specified field or fields. Each meaningful word or term is listed in alphabetical order and certain word frequency statistics are collected such as number of content words in list, number of trivial words (e.g., a, an, the), average word length, and distribution of word lengths. VOCAB is useful as a tool in assigning index terms or in constructing search formulas since the vocabulary lists indicate the terms actually used in the citations.

SEARCH scans the user's collection and retrieves those citations which match the search formula.

Boolean logic is used to combine the terms or words in as inclusive or specific a search formula as wished. Multiple searches can be conducted by following the first search formula with a second, and so on.

Notes

[1] FAMULUS is not an acronym. A Famulus was the private secretary of a medieval scholar or alchemist. Perhaps the most well-known was Wagner, Famulus of Dr. Faustus.

[2] Numbers refer to Appendix I, where a short summary of each collection is given.

[3] Parker, E. B. *The user's place in an information system.* Amer. Doc. 17 (1) : 26. 1966.

References

Cuadra, C. A.
1967. *Progress in information science and technology.* FID & Int. Congr. in Doc. III b1, 1-16.

Engelbart, D. C.
1961. *Special considerations of the individual as a user, generator, and retriever of information.* Amer. Doc. 12(2) : 121-125.

Herner, S.
1958. *The relationship of information user studies and the design of information storage and retrieval systems.* Air Res. & Dev. Command, Griffis Air Force Base, New York, (AD 213 781) Tech. Note 1. 24 p.

Maisell, R. E.
1960. *Information gathering patterns and creativity: a study of research chemists in an industrial research laboratory.* Amer. Doc. 11(1) : 9-17.

Menzel, H.
1966. *Can scientific information needs be ascertained empirically?* Proc. 2nd Int. Symp. on Communication Theory and Res.: 279-294. Washington, D. C.

Yerke, T. B.
1966. *Adapting library computer programs to individual documentation.* Proc. 2nd Annual Amer. Water Resources Conf.: 432-440. Chicago, Ill.

almost impossible. When the input and output are the responsibility and need of one person, the relevancy problem is telescoped considerably. If results of retrieval attempts are unsatisfactory, a fault in system design may be involved, or it may be that the user has not sufficiently analyzed his information needs and is not monitoring the information sources which would assure the optimum input into the system.

Much remains to be learned about the benefits which FAMULUS can bring to a group of researchers in diverse fields working within a large research institution. Currently on-going research is aimed at analyzing what types of benefits these users are receiving and how an optimum interface between personal and formal systems can be developed.

APPENDIX I

The structure and content of literature collections on FAMULUS cited in this paper:

(1) User : H. W. Anderson, U.S. Forest Service, Berkeley, Calif.
Subject : Analytical Methods and Research in Hydrology and Sedimentation.
Size : 3,000 citations averaging 2,500 characters (40 EAM cards) in length.
Fields : Author, title, collation, series note, keywords, abstracts, comments, references.
Abstracts and keywords are developed by Anderson.

(2) User : Professor P. S. Messenger, University of California, Berkeley.
Subject : Ecological Entomology.
Size : 3,300 citations averaging 400 characters (seven EAM cards) in length.
Fields : Author, title, date of publication, source of publication, keywords, accession number.
Collection is heavily indexed.

(3) User : Remote Sensing Project, U.S. Forest Service, Berkeley, Calif.
Subject : Remote Sensing of the Environment (i.e., detection of environmental conditions via sensors such as aerial photography, etc.)
Size : 2,300 citations averaging 400 characters (seven EAM cards) in length.
Fields : Author, title, collation, major category classification, sub-category classification, accession number.
Categories were developed by the various project members.

(4) User : D. A. Sharpnack, P. G. Langley, U. S. Forest Service, Berkeley, Calif.
Subject : Mensuration Techniques in Forestry.
Size : 1,300 citations averaging 250 characters in length.
Fields : Author, title, collation, keywords, accession number.

(5) User : Remote Sensing Laboratory, School of Forestry and Conservation, University of California, Berkeley.
Subject : Remote Sensing of the Non-Urban Environment.
Size : 3,500 citations averaging 250 characters (five EAM cards) in length.
Fields : Author, title, collation, keywords, location of item, comments.

(6) User : Mark J. Schroeder, U.S. Forest Service, Fort Collins, Colorado.
Subject : Fire Danger Rating (includes research on fire-oriented climatology, meteorology, etc.)
Size : This is a new collection and currently has about 500 citations in it, averaging about 700 characters.
Fields : Author, title, publisher, publication date, keywords, location of item, abstract, accession number.

(7) User : C. E. Conrad, U.S. Forest Service, Berkeley, Calif.
Subject : Range Management Techniques.
Size : 500 citations averaging 700 characters (ten EAM cards) in length.
Fields: : Author (including author's affiliation), title, keywords, citation information and location, kind of paper (report, review, bibliography, etc.), objective of paper, analysis of data in paper, evaluation, comments.

(8) User : C. E. Crisp, U.S. Forest Service, Berkeley, Calif.
Subject : Plant Pathology.
Size : 500 citations averaging about 200 characters (six EAM cards) in length.
Fields: Author, title, date of publication, publication, keywords, accession number, notes.

FAMULUS leaves him free to index individually and creatively, according to his needs. At the same time, there are always common features between files so that intercollection searching or listing is practicable. FAMULUS plays a role here of educator, as well as retrieval tool, for were it not for the facility with which bibliographies and search queries can be exchanged between users, this activity would not take place at the level it presently does.

THE FAMULUS SYSTEM

The FAMULUS system consists of eight sub-systems, each with a variety of functions. The sub-systems are named after their principal function: EDIT detects format errors and produces the basic input tape; SORT puts the file in order by any requested field (e.g., author, title, location, series, etc.); MERGE updates files; INDEX shows what citations file behind each retrieval term; VOCAB constructs retrieval vocabularies from title words or terms added by the user. There can be two sets of terms, i.e., "free" terms from titles or controlled terms added by the scientists; SEARCH does the retrieval work, operating on Boolean logic; GALLEY produces printouts in several formats; OSSIFY produces safety decks from master tapes.

All FAMULUS-produced tapes for a given computer are compatible. This permits scientists to duplicate their files or portions of them for loan to co-workers. Files within and outside of projects can also be merged and compared. A minimum of control cards is needed to operate the programs. Scientists and their clerks can learn how to set up a computer run with very little effort. Less than 10 cards will instruct the system to mount the proper magnetic tapes and process the collection as desired. When new citations on card decks are being input, the input deck must, of course, be added to the control cards.

FAMULUS users have been satisfied with the costs of using the system, which they pay out of their project budgets. Useage requires about 4.5 to 5 minutes of CDC 6400 computer time to edit a collection of 500 citations, write them on tape, sort them as the user wants, index them, and provide the user with a printout in the format option he has chosen. The dollar cost will vary among computer centers, but the time figure is relatively constant for the same model computer and the reader can calculate the cost in his environment. Cost does not increase linearly with increase in citations. To process 1,000 citations as above will take about 7 minutes rather than 10.

The figures given are for "standard" citations which average seven EAM cards. If a collection has only 500 citations but each is 40 EAM cards in length (due to long abstracts), the handling bulk will be greater than for a collection of 1,000 citations averaging 10 cards each. The sorting and indexing of the longer citations will take no more time, but editing and printing will. FAMULUS handles any configuration with great efficiency; a tape with 10,000 "standard" citations can be searched in 2 minutes. FAMULUS shows that this kind of processing on larger computers is quite inexpensive.

The basic cost is that of keypunching already-existing reference files. In the case of new collections, we have shown potential users that the cost of keypunching in the FAMULUS input format is very little more than that of typing. Nonetheless, a large backlog is an expensive and time-consuming problem. It may be confronted head-on, through the hiring of a clerk who will spend her major time at the keypunch. Not all our projects or scientists are this opulent. The usual procedure is to feed small batches into the keypunch pool over a period of time. In 18 months one scientist caught up to a backlog of 3,000 items. He now adds some 50 citations per month.

FAMULUS may be used for applications other than the handling of personal literature files. Any input material which can be keypunched and subsequently manipulated meaningfully by FAMULUS is adaptable. Five such special applications have been found during the first year of operation. These include processing the results of a job attitude survey conducted in the California State Division of Forestry by the University of California; preparation of an on-your-desk, constantly revised personnel information file of employees at the Pacific Southwest Station for use in the personnel office; a survey of library-use by the Station personnel, including interlibrary loans; an inventory of computer facilities available to all Forest Service Experiment Stations in the United States; and a catalog of the computer programs held at this Station. FAMULUS also produces the Library's author catalog and title catalog as well as the monthly accessions list, which is both indexed and classified.

Our present FAMULUS users report very satisfactory retrieval relevance. All information systems are source controlled,[3] and nothing can be gotten out of them which was not first put in. When the input to a system is controlled by one set of persons and the output is desired by another (as obtains between indexing services and their clientele), at least two sets of relevancy judgment are involved, and congruity is

The Authors

HILARY D. BURTON has been computer services librarian of the Station since 1965. Native of Pittsburgh, Pa., she earned a A.B. degree (1965) in Political Science and an M.L.S. (1966) degree at the University of California, Berkeley. **ROBERT M. RUSSELL,** native of Detroit, Mich., has been a computer programer at the Station since 1966. He was graduated from Graceland College in 1953, and also holds a B.S. (1956) degree in mathematics from the University of Michigan. **THEODOR B. YERKE** became Station Librarian in 1960. Born in Great Falls, Mont., he earned bachelor's degrees at the University of California, Los Angeles (1949, German language and literature) and the University of California, Berkeley (1952, library science).

The Forest Service of the U.S. Department of Agriculture

. . . Conducts forest and range research at more than 75 locations from Puerto Rico to Alaska and Hawaii.

. . . Participates with all State forestry agencies in cooperative programs to protect and improve the Nation's 395 million acres of State, local, and private forest lands.

. . . Manages and protects the 187-million-acre National Forest System for sustained yield of its many products and services.

The Pacific Southwest Forest and Range Experiment Station

represents the research branch of the Forest Service in California and Hawaii.

GPO-977-577